AUTOMATED TASK CLASSIFICATION SYSTEM

Cross Reference To Related Application

This application is a continuation of U.S. Patent Application Serial Number 08/943,944, which is a continuation-in-part of U.S. Patent No. 5,675,707. U.S. Patent Application Serial Number 08/943,944 and U.S. Patent No. 5,675,707 are hereby incorporated herein by reference in their entireties.

Technical Field

This invention relates to verbal and non-verbal speech processing, and more particularly to an automated task classification system related to the performance of one or more desired tasks.

Background Of The Invention

In telephonic communications it is well known for the recipient of a call, particularly in a commercial setting, to utilize an automated call routing system which initially presents the calling party with a menu of routing choices from which the caller is asked to select by pressing particular numbers on the keypad associated with the caller's telephone -- the routing system recognizing a tone associated with the key pressed by the caller. It is also known to offer such callers a choice between pressing a number on the keypad to select a menu choice or to say that number -- e.g., "press or say one for customer service". In the particular context of telephone services, it is also known to use an automatic routing system for selecting among billing options for a call placed over such a service. For example, in the case of a long distance telephone call to be billed to other than the originating number, a menu may be presented to the calling party in the form, "please say 'collect', 'calling card', 'third number' or 'operator'".

While such routing systems work reasonably well in cases where the number of routing choices is small, if the number of selections exceeds about 4 or 5, multi-tiered

menus generally become necessary. Such multi-tiered menus are very unpopular with callers -- from the perspective of a typical caller, the time and effort required to navigate through several menu layers to reach a desired objective can seem interminable. Equally important, from the perspective of both the caller and the recipient, the percentage of successful routings though such a multi-tiered menu structure can be quite low, in some cases, less than 40 percent. Stated differently, in such circumstances, more than 60 percent of the calls accessing such a multi-tiered menu structure might be either terminated without the caller having reached the desired objective or else defaulted to an operator (or other manned default station).

To address these limitations in the prior art, it would be desirable to provide a system which can understand and act upon verbal and non-verbal input from people. Traditionally, in such speech understanding systems, meaningful words, phrases and structures have been manually constructed, involving much labor and leading to fragile systems which are not robust in real environments. A major objective, therefore, would be a speech understanding system which is trainable, adaptive and robust -- i.e., a system for automatically learning the language for its task.

Summary Of Invention

The invention concerns an automated task classification system that operates on a task objective of a user. The system may include a meaningful phrase generator that generates a plurality of meaningful phrases from a set of verbal and non-verbal speech. Each of the meaningful phrases may be generated based on one of a predetermined set of the task objectives. A recognizer may recognize at least one of the generated meaningful phrases in an input communication of the user and a task classifier may make a classification decision in response to the recognized meaningful phrases relating to one of the set of predetermined task objectives.

Brief Description Of The Drawings

FIG. 1 provides illustrative examples of false and missed detection by a classifier for an automated call routing system based on use of "meaningful phrases".

FIG. 2 provides illustrative examples of correct detection by a classifier for an

automated call routing system based on use of "meaningful phrases".

FIG. 3 depicts an illustrative example of the advantage provided by the "meaningful phrase" classification parameter of the system of the invention.

FIG. 4 presents in block diagram form the structure of the system of the invention.

FIG. 5 depicts the methodology used by the system of the invention in flowchart form.

FIG. 6 provides illustrative examples of "meaningful phrases" determined according to the invention.

Detailed Description of the Preferred Embodiments

The discussion following will be presented partly in terms of algorithms and symbolic representations of operations on data bits within a computer system. As will be understood, these algorithmic descriptions and representations are a means ordinarily used by those skilled in the computer processing arts to convey the substance of their work to others skilled in the art.

As used herein (and generally) an algorithm may be seen as a self-contained sequence of steps leading to a desired result. These steps generally involve manipulations of physical quantities. Usually, though not necessarily, these quantities take the form of electrical or magnetic signals capable of being stored, transferred, combined, compared and otherwise manipulated. For convenience of reference, as well as to comport with common usage, these signals will be described from time to time in terms of bits, values, elements, symbols, characters, terms, numbers, or the like. However, it should be emphasized that these and similar terms are to be associated with appropriate physical quantities B such terms being merely convenient labels applied to those quantities.

It is important as well that the distinction between the system's operations and operating a computer, and the system's computation itself should be kept in mind. The present invention relates to a system that operates a computer in processing electrical or other (e.g., mechanical, chemical physical signals to generate other desired physical

signals.

For clarity of explanation, the illustrative embodiment of the present invention is presented as comprising individual functional blocks (including functional blocks labeled as "processors"). The function these blocks represent may be provided through the use of either shared or dedicated hardware, including, but not limited to hardware capable of executing software. For example the functions of processors presented in Figure 4 may be provided by a single shared processor. (Use of the term "processor" should not be construed to refer exclusively to hardware capable of executing software.)

Illustrative embodiments may comprise microprocessor and/or digital signal processor (DSP) hardware, such as the AT&T DSP16 or DSP32C, read-only memory (ROM) for storing software performing the operations discussed below, and random access memory (RAM) for storing results. Very large scale integration (VLSI) hardware embodiments, as well as custom VLSI circuitry in combination with a general purpose DSP circuit, may also be provided.

A fundamental aspect of the invention is a task classification and routing service that shifts the burden of understanding a specialized vocabulary from the initiating party to the recipient. Thus, while the task classification system of the invention may be applied to other disciplines and technologies using linguistic and visual sensory inputs, for ease of discussion, a generalized embodiment of the invention is represented as a call routing system having the following characteristics:

First, a caller accessing the system will be presented with a greeting similar to "How may I help you?"

After a caller responds to that greeting with a natural speech statement of the caller's objective (such as a routing objective), the system is able to either classify the caller's request into one of a number of predefined objective routings and implement that routing, or to transfer the caller to an operator where the request did not fit one of the predefined objective routings or the system was unable to understand the caller's request.

To minimize erroneous routings, as well as transfers to an operator

position when a caller actually wanted one of the predefined routing objectives, the system also incorporates a dialog function to obtain additional information from the caller for improving the system's confidence in a classification of a caller's objective.

Hereafter, an embodiment of the invention will be described based on access by a caller to a telephone service provider. In such an embodiment, routing objectives for a caller may include call billing options (e.g., collect, third-party), dialing information, billing questions, credit requests (as for a wrong or mis-dialed number), area codes, etc.

In traditional telephone switching systems, a user is often required to know separate numbers and/or dialing patterns to access different services provided by a telephone carrier, as well as possibly having to navigate a menu-driven system which then routes the user to the desired objective. In the system of the invention the user is able to access a central number and the user's objective will be implemented by the telephone carrier on the basis of its content.

An example of such content-based routing would be where a caller responds to a "How may I help you?" prompt with "I want to *reverse the charges", whence* the appropriate action is to connect the caller to an automated subsystem which processes collect calls. Another example would be a caller response of "I am having a *problem understanding my bill",* in which case the caller should be connected to the telephone carrier's business office. The system thus needs to understand spoken language to the extent of routing the call appropriately.

The system of the invention concerns the needs to understand language, both verbal and non-verbal, to the extent that a user's task objectives may be implemented. Earlier work describing the understanding of multi-sensory (or verbal and non-verbal) language is found in Gorin, A., "On automated language acquisition", J. Acoust. Soc. Am., 97 3441-3461, (June 1995) [hereafter referred to as *Gorin 95*], which is incorporated herein and made a part hereof. In this paper, the system described receives some input, comprising linguistic and possibly other stimuli, denoted as "language". In response to this input, it performs some action. See p. 3441.

The paper also discusses the construction of *sensory primitive subnetworks*, which learn cross-channel (i.e. multimodal) associations between different stimuli. Experiments discussed involve machines that receive both linguistic and visual input on respective channels, such as a focus of attention on a particular object. In this regard, a structure is described in the environment, in that input signals to the system can be organized into symbols and can be imposed on all sensory inputs. See page 3443.

Thus, in many situations of interest, the appropriate machine response depends not only on the spoken input but upon the state of the environment. In other words, the invention described below concerns devices with multisensory inputs (linguistic and visual) including verbal, non-verbal, and multimodal inputs. See p. 3450.

A number of considerations from that baseline approach are material to the system of the invention. Certain of those considerations will be briefly reviewed hereafter.

Central to the approach here is a database of a large number of utterances, each of which is related to one of a predetermined set of routing objectives. This database forms an input to a classification parameter algorithm. Preferably, such utterances will be extracted from actual user responses to a prompt of "How may I help you?" (or similar words to the same effect). Each utterance is then transcribed and labeled with one of the predetermined set of routing objectives. Illustrative utterances from the database utilized by the inventors are as follows:

Yeah, I want to reverse the charges

I was just disconnected from this number

I was trying to hang-up

I am trying reach Mexico

Charge this to my home phone

In a related article co-authored by one of the inventors [Gorin, A. L. , Hanek, H., Rose, R. and Miller, L., "Spoken Language Acquisition for Automated Call Routing", *in Proceedings of the International Conference on Spoken Language Processing (ICSLP 94),* Yokohama (Sept. 18-22, 1994)] [hereafter *Gorin 94A*], it is noted that the

distribution of routing objectives in such a database may be substantially skewed. The implications of such skewing may well be taken into account in the determination of the particular set of routing objectives to be supported on an automated basis by the system of the invention.

A *salience* principle as related to the system of the invention has been defined in another article co-authored by one of the inventors [Gorin, A.L., Levinson, S.E. and Sankar, A. "An Experiment in Spoken Language Acquisition," *IEEE Trans. on Speech and Audio, Vol. 2,* No. 1, Part II, pp. 224-240 (Jan. 1994)] [hereafter *Gorin 94*]. Specifically, the *salience* of a word is defined as *the information content of that word for the task under consideration.* It can be interpreted as a measure of how meaningful that word is for the task. Salience can be distinguished from and compared to the traditional Shannon information content, which measures the uncertainty that a word will occur. As is known, such traditional information content can be estimated from examples of the language, while an estimation of salience requires both language and its extra-linguistic associations.

As previously noted, the baseline approach of *Gorin* 95 (which has been incorporated herein by reference) uses as a classification parameter words from test speech utterances which have a salient association with particular objective routings. A significant point of departure for the system's methodology of the invention from that baseline approach resides in the use of *meaningful phrases* as the classification parameter, a clear improvement over that baseline approach. Before describing the nature of that improvement, or the system's methodology for determining such *meaningful phrases,* it is useful to define two types of error experienced in such an automated call routing system and a related "success" concept:

False detection of a routing objective occurs when a salient (meaningful) phrase related to one routing objective is detected in a caller's input speech when the caller's actual request was directed to another routing objective. The probability of such a false detection occurring will hereafter be referred to by the designation: P_{FD}

Missed detection of a routing objective occurs when the callers input speech is directed to that routing objective and none of the meaningful phrases which are associated with that routing objective are detected in the input speech. The probability of such a missed detection occurring will hereafter be referred to by the designation: P_{MD}

Coverage for a routing objective refers to the number of successful translations by the system of a request for a routing objective to that routing objective relative to the total number of input requests for that routing objective. As an illustrative example, a routing objective for which 60 successful translations occurred out of 100 input requests for that routing objective would be said to experience 60% coverage. It is noted that *Coverage* = $1-P_{MD}$

Of the two error types defined above, one is significantly more "costly" than the other in an automated call router. The consequence of a *false detection* error is the routing of a caller to a different routing objective than was requested by the caller. Such a result is at least very annoying to the caller. The possibility also exists that such an error could result in a direct cost to the system provider -- an annoyed customer or potential customer being classified here as an indirect cost -- through some non-system error resulting from the caller being connected to an incorrect routing objective. The consequence of a *missed detection* error, on the other hand, is simply the routing of the caller to a default operator position and the only cost is the lost opportunity cost of not handling that particular call on an automated basis. Thus, while ideally the probabilities of both *missed detection* and *false detection* would be near zero, it is far more important that this objective be realized for *false detection* errors. As will be seen below, there are circumstances where tradeoffs must be made between minimizing one or another of these error probabilities, and this principle will be applied in such circumstances.

Figure 1 provides several illustrative examples of False Detections and Missed Detections from the database of speech utterances used by the inventors. While the basis for error in each of these examples is believed to be largely self-explanatory, the error in the first example in each set will be briefly described. In the first example under

False Detection, the meaningful phrase is I NEED CREDIT, and thus this phrase would have been classified as a request for credit. However, from reading the entire utterance, it is apparent that the caller actually wanted to be transferred to another carrier (the carrier receiving this request being AT&T). In the first example under Missed Detections, there are no meaningful phrases identified in the utterance (and therefore no basis for classifying the caller's objective), although it is apparent from reading the utterance that the caller is seeking a billing credit. As a comparative illustration, Figure 2 shows several examples of correct detection of a billing credit objective from meaningful phrases in the input speech.

There are two significant advantages of the system's methodology of the invention in using *meaningful phrases as* the classification parameter over the use of *salient* words in the baseline approach described in *Gorin 95*. First, with the use of words as the classification parameter, the word choices for detecting a given routing objective may be highly limited in order to achieve a minimum probability of *false detection* -- i.e. use of only words having a near likelihood of predicting the intended routing objective – and therefore the *coverage* for such a routing objective is likely to be very low, leading to a high probability of *missed detection* errors occurring. With *meaningful phrases* as a classification parameter, on the other hand, both low probability of *false detection* and low probability of *missed detection* are achievable.

Figure 3 provides an illustrative example of this advantage. That figure shows the Classification Rate and the Coverage for an exemplary routing objective, Billing Credit, as the phrase used for the classification parameter grows in length and/or complexity. The Classification Rate is defined as the probability of the routing objective (CREDIT) having been requested, given the occurrence of the selected phrase in the input speech (i.e., P(CREDIT I phrase). Similarly, the Coverage term is defined as the probability of the selected phrase appearing in the input speech, given that the designated routing objective (CREDIT) has been requested. In the Phrase column, parenthesis surrounding a series of terms separated by " I " indicate one of those terms appearing in the indicated position with other terms in that row. The nomenclature

"F(Wrong)" indicates a grammar fragment surrounding the word "wrong", the phrase in the fourth row of that column being representative of such a grammar fragment surrounding a salient word. The designation "previous" indicates a carry forward of everything on the preceding line. And finally, the abbreviation "eos" indicates "end of sentence".

The second area of improvement relates to the speech recognition function which operates on the caller's input speech. Essentially, at the present state of the art of speech recognition systems, the larger the fragment of speech presented to such a speech recognizer, the higher the probability of a correct recognition of that speech fragment. Thus, a speech recognizer programmed to spot one of a set of salient words can be expected to fail in its task significantly more often than such a device programmed to spot meaningful phrases, comprising two or more words.

Figure 4 shows in block diagram form an exemplary structure that may perform the inventive system's method. As can be seen from the figure, that structure comprises two related subsystems: Meaningful phrase generation subsystem 1 and Input speech classification subsystem 2. As already described, Meaningful phrase generation subsystem 1 operates on a database of a large number of utterances each of which is related to one of a predetermined set of routing objectives, where each such utterance is labeled with its associated routing objective. The operation of this subsystem is essentially carried out by Meaningful phrase processor 10 which produces as an output a set of meaningful phrases having a probabalistic relationship with one or more of the set of predetermined routing objectives with which the input speech utterances are associated. The operation of Meaningful phrase processor 10 is generally determined in accordance with a grammatical inference algorithm, described below.

Operation of Input speech classification subsystem 2 begins with the inputting of a caller's routing objective request, in the caller's natural speech, to Input Speech Recognizer 15. That Input Speech Recognizer may be of any known design and performs the function of detecting, or spotting, the existence of one or more meaningful

phrases in the input speech. As can be seen in the figure, the meaningful phrases developed by Meaningful phrase generation subsystem 1 are provided as an input to Speech Recognizer 15.

The output of Speech Recognizer 15, which will comprise the recognized meaningful phrase(s) appearing in the caller's routing objective request, is provided to Interpretation Module 20. That Interpretation Module applies a confidence function, based on the probabalistic relation between the recognized meaningful phrase(s) and selected routing objectives, and makes a decision either to implement the chosen routing objective (in the case of high confidence) with an announcement to the caller that such an objective is being implemented, or to seek additional information and/or confirmation from the caller via User Dialog Module 25, in the case of lower confidence levels. Based on caller feedback in response to such an inquiry generated by Interpretation Module 20, the Interpretation Module again applies the confidence function to determine if a decision to implement a particular routing objective is appropriate. This feedback process continues until either a decision can be made to implement a particular routing objective or a determination is made that no- such decision is likely, in which case the caller is defaulted to an operator position.

As will thus be apparent, the meaningful phrases developed by Meaningful phrase generation subsystem 1 are used by Input Speech Recognizer 15, to define the phrases which the Recognizer is programmed to spot, and by Interpretation Module 20, both to define the routing objectives related to meaningful phrases input from Speech Recognizer 15 and to establish the level of confidence for a relation of such input meaningful phrase(s) to a particular routing objective.

The system's methodology of the invention is graphically illustrated in the flowchart of Figure 5. It will be appreciated from prior discussion that the process depicted in Figure 5 is broadly separated between Meaningful Phrase Generation functions 100 and Automated Call Routing functions 200. Considering first the Meaningful Phrase Generation functions, a database of speech utterances labeled with a requested objective is provided at step 105, and that database is accessed at step

110 to extract a set of n-grams. The mutual information of those n-grams is determined at step 115 and the MI values so determined are compared with a predetermined threshold at step 120. n-grams having an MI value below the threshold are discarded and those above the threshold are operated on by salience measuring step 125. In step 130 those salience values are compared with a predetermined salience threshold and those n-grams having a salience value below the threshold are discarded. n-grams passing the salience threshold test are stored at step 135 as the set *of meaningful phrases.*

The Automated Call Routing functions begin with step 205 where a caller accessing the router is provided a greeting similar to "How may I help you?" The caller's response to that greeting is obtained at step 210 and that response is examined for the presence *of* one or more *meaningful phrases. As* can be seen in the figure, the set of *meaningful phrases* will have been provided as an input to step 215 as well as step 220. In step 220, an attempt is made to classify the caller's objective based on *meaningful phrase(s)* found in the caller's input speech. A confidence factor is invoked at step 225, and a decision is made as to whether a sufficient confidence level exist to implement the classified objective. If yes, the objective is implemented, at step 245. If the confidence function dictates either that an affirmation from the user or more information is needed, a dialog with the user is carried out at step 230. Such dialog will typically begin with a query to the user of the form "You want ... ?" If the user responds in the negative to such a query, at step 235, the process returns to step 210 with a restatement by the user of its request. A "yes" or silence response from the user moves the process along to step 240, where consideration is given to whether other information is needed to carry out the objective which was not in the input speech -- e.g., the number to be credited in the case of a credit request. Where additional information is needed, the process returns to step 230 for additional dialog with the user. If no additional information is required, the objective is implemented at step 245.

As will be understood at this point, a fundamental focus of this invention is that of providing a system which learns to understand and act upon spoken input. It will be

apparent that the ultimate goal of such a speech understanding system will be to extract meaning from the speech signal. Moreover, as was shown in Gorin 95, for systems which understand linguistic and visual language, the *semantic* aspects of communications are highly significant. Not surprisingly, such semantic considerations are fundamental to the development of *the meaningful phrases* used as classification parameters in the system of the invention.

The determination of the *meaningful phrases* used by the invention is founded in the concept of combining a measure of commonality of words and/or structure within the language -- i.e., how often groupings of things co-occur -- with a measure of significance to a defined task for such a grouping. In the preferred embodiment of the invention, that commonality measure within the language is manifested as the *mutual information* in n-grams derived from a database of training speech utterances and the measure of usefulness to a task is manifested as a salience measure. Other manifestations of these general concepts will be apparent to those skilled in the art.

As is known, mutual information ("MI"), which measures the likelihood of co-occurrence for two or more words, involves only the language itself. For example, given *War and Peace* in the original Russian, one could compute the mutual information for all the possible pairings of words in that text without ever understanding a word of the language in which it is written. In contrast, computing salience involves both the language and its extra-linguistic associations to a device's environment. Through the use of such a combination of MI and a salience factor, phrases may be selected which have both a positive MI (indicating relative strong association among the words comprising the phrase) and a high salience value.

The *meaningful phrase* determination is implemented in a Grammatical Inference ("GI") algorithm. That GI algorithm searches the set of phrases which occur in the training database using as a selection criteria both the mutual information (within the language) of a phrase and a measure of the salience of the phrase for a task under consideration. In its general form the GI algorithm carries out the following steps.

1. n = number of words in a phrase to be evaluated.

2. Set n = 1.
3. Generate a list of phrases on length n in training database.
4. Evaluate commonality of those phrases (as function of frequency/probability) and salience of the phrases.
5. Select subset according to a predetermined threshold.
6. Generate list of phrases of length n+1 by expanding the set of phrases on length n.
7. Evaluate mutual information and salience for the phrases generated in step 6.
8. Select subset according to a predetermined threshold.
9. Set n=n+1
10. Go to step 6.

The algorithm may be carried out to any level of n desired and the selection threshold may be varied up or down with successive iterations to vary the output of the algorithm. That output will be a set of meaningful phrases.

An exemplary output of this algorithm is shown in Figure 7. In that figure, the exemplary meaningful phrases are shown along with the MI value, the salience value, and the Action to which the meaningful phrase is directed at the probability value shown. Note that the Max Probability (sometimes designated P_{max} is used as a proxy for the salience value. That Max Probability is defined as the maximum of the distribution of probabilities of each action, given the phrase under consideration, which probabilities are determined by a sub-step of the GI algorithm.

As an aid to further understanding of the invention, two illustrative dialogues from the operation of the system of the invention are hereafter presented. In Session One, the system immediately understands the input, based on learned associations between the content of the sentence and the selected routing objective of *third-number billing.*

Hello, how may I help you?

"I want to charge this call to my home phone."

You want to bill this call to a third party.

Please enter the billing number:

As can be seen, the absence of a response from the caller during the pause between the second and third inquiries from the system was interpreted as assent to the second inquiry.

In Session Two, the system encounters an ambiguous input, which is resolved via dialog. Note the hesitant response, based on the system's confidence model. Observe also that the second input involves both negative reinforcement *(no)* and additional clarifying information.

Hello, how may I help you?

"Can I charge this call please?"

Do you want to bill this call to a third party?

"No, I want to put this on my Universal card."

You want to bill this call to a credit card.

Please enter the card number:

A novel automated call routing system has been disclosed which carries out the function of searching for a classification parameter in natural speech, that classification parameter being manifested as a set of meaningful phrases, which are themselves defined in terms of mutual information and a salience factor. Then, depending on whether one or more particular meaningful phrases are found in the input speech, a decision rule is applied to classify the appropriate routing objective. An important advantage of the system of the invention is that the vocabulary and grammar for the system are unconstrained, being acquired by the system during the course of performing its task. By contrast, in prior art systems, the salient vocabulary words and their meanings are explicitly provided to the system.

While prior art language recognition systems have been very much handcrafted and lacking in robustness, the system of the invention brings an automated training procedure to the language recognition function -- automated procedures for determining *meaningful phrases* in test speech utterances marked with associated objective routings B and then finds such *meaningful phrases* in caller input speech (possibly including confidence enhancing dialog with the caller) and implements the caller-requested objective routing.

Although the present embodiment of the invention has been described in detail, it should be understood that various changes, alterations and substitutions can be made therein without departing from the spirit and scope of the invention as defined by the appended claims.